RANDOM EFFECTS MODELS FOR BINARY DATA APPLIED TO ENVIRONMENTAL/ECOLOGICAL STUDIES

by

Charles E. McCulloch
Biometrics Unit and Statistics Center
Cornell University

BU-1120-M April 1991

ABSTRACT

The outcome measure in many environmental and ecological studies is binary, which complicates the use of random effects. The lack of methods for such data is due in a large part to both the difficulty of specifying realistic models, and once specified, to their computational intractibility.

In this paper we consider a number of examples and review some of the approaches that have been proposed to deal with random effects in binary data. We consider models for random effects in binary data and analysis and computation strategies. We focus in particular on probit-normal and logit-normal models and on a data set quantifying the reproductive success in aphids.

Random Effects Models for Binary Data Applied to Environmental/Ecological Studies

by Charles E. McCulloch

1. INTRODUCTION

The outcome measure in many environmental and ecological studies is binary, which complicates the use of random effects since such models are much less well developed than for continuous data. The lack of methods for such data is due in a large part to both the difficulty of specifying realistic models, and once specified, to their computational intractibility.

In this paper we consider a number of examples and review some of the approaches that have been proposed to deal with random effects in binary data. The advantages and disadvantages of each are discussed. We focus in particular on a data set quantifying the reproductive success in aphids.

2. EXAMPLES

In this section we consider several examples in order to motivate later discussion.

Example 1: Reproductive success in aphids

Female aphids are gathered from the field early in the summer and late in the summer. Each aphid is used to raise a clonal line. The dependent variable is reproductive success (does the insect survive to reproduce?), recorded for each of several individuals from each line. Interest centers primarily on whether there is variation from clone to clone and secondarily on the fixed effects of time (early versus late), the crop on which the individual is raised (alfalfa or clover) and plot (plot from which the alfalfa or clover was taken).

We will describe this data set in some detail since it is used to illustrate some of the models and fitting techniques later (data courtesy of Professor S. Via at Cornell University). 28 female aphids were collected from the field in both the early and late summer. Clonal lines were raised from each female in the laboratory in two separate chambers (sublines). For each clonal subline 1 to 4 females were raised on alfalfa and on clover which came from plot 2 or plot 2. If the data were balanced there would be 2 times $\times$ 28 clones $\times$ 2 sublines $\times$ 2 crops $\times$ 2 plots $=$ 448 observations. In actuality there were 412 individuals tested. Each individual was recorded as surviving to reproduce or not.

So there are two random effects (clone and subline) and five fixed effects (constant term,

crop, time, plot, and plot by crop interaction).

Example 2: Salamander mating (taken from McCullagh and Nelder, 1989).

Three experiments were conducted in which 10 males and 10 females from each of two populations were cross-bred. Comparison of the probabilities of mating within and between populations was of interest, as well as the variability in male and female mating probabilities.

Example 3: Relation of asthma attacks to pollution level (taken from Stiratelli, Laird and Ware, 1984).

Sixty-four asthmatics are studied for over 200 consecutive days. The dependent variable was the occurrence of an asthma attack on a particular day. Independent variables were pollution and weather variables, occurence of an asthma attack on the previous day and individual specific variables (sex, age, etc.). Random effects included individual effects and individual specific parameters for the pollution variable.

Interest focusses both on the effect of the pollutants and on estimates of each individual's response to the pollutants, i.e., are some individuals more sensitive?

Example 4: Teratology experiments with litter effects (taken from Weil, 1970).

Sixteen pregnant rats recieved a control diet and 16 received a chemically treated diet to test for carcinogenity. Each offspring alive after four days was followed to day 21 and survival was recorded. So the fixed effect is diet (control versus treated) and the random effect is litter. Interest is solely in the fixed effect and the random effect is a nuisance factor.

Of course, each of these examples has a binary dependent variable and random effects. However, they differ greatly in their inferential goals. In example 1, primary interest is in estimating the variance of the random effect due to clone. In examples 2 and 3, both the fixed effect and the random effects are of interest. In example 3, interest focusses on a Bayes or empirical Bayes analysis in order to derive predictions of the values of the random effect. In example 4, the random effects are merely nuisance and interest is in estimating and testing hypotheses about the fixed effect.

3. WHY BINARY DATA MODELS ARE DIFFERENT

To see why models which incorporate random effects are more difficult to specify for binary data, consider first the construction of models for continuous data. Let y denote the data vector and u the random effects (not including the error term). The error term is defined by attributing a distribution to y-E[y|u], usually $\mathcal{N}(0, \sigma_e^2)$, and in almost every case, a distribution having constant variance, independent of the value of the mean of y or the conditional mean of y. This is not a reasonable assumption for binary data. For binary data

y_i is distributed as a Bernoulli random variable with probability of success $p_i = P\{y_i = 1\} = E[y_i]$ and variance $E[y_i][1 - E(y_i)]$. As the mean of y_i approaches one or zero, the variance approaches zero and this dependence between mean and variance must be included in any reasonable model. Thus a model with an additive error component with fixed variance is inadequate.

Further problems arise when specifying the distribution of random effects. For simplicity, consider a model for a binary variable, y_{ij}, with a single fixed effect, βx_{ij}, and a single random effect, α_i. Conditional on the random effects, the mean of y_{ij} will be taken as

$$E(y_{ij} \mid \alpha_i) = \beta x_{ij} + \alpha_i \ . \tag{1}$$

For the continuous data situation the α_i are usually assumed to be *i.i.d* with variance σ_α^2 and are often assumed to have a normal distribution. For the binary data situation, since the mean or conditional mean of y_{ij} cannot be larger than one or less than zero, the α_i cannot have a normal distribution, and as the mean of y_{ij} approaches zero or one the variance of the α_i must approach zero. So the distribution of the α_i also cannot have a fixed variance. The usual way of accommodating these requirements is to consider nonlinear models which allow the random effects to enter into the conditional mean in a non-additive fashion.

A common model for binary data is the logistic regression model where y_{ij} has a Bernoulli distribution with probability of success of p_{ij} and logit(p_{ij}), defined as logit(p_{ij}) = $\log[p_{ij}/(1-p_{ij})]$, is assumed to be linear in the effects. A mixed model analogous to (1) could be defined as:

$$\begin{gathered} \text{Conditional on the } \alpha_i,\ y_{ij} \sim \text{Bernoulli}(E[y_{ij}|\alpha_i]), \text{ where} \\ \text{logit}(E[y_{ij}|\alpha_i]) = \beta x_{ij} + \alpha_i \quad \text{and} \quad \alpha_i \sim \textit{i.i.d.}\ \mathcal{N}(0, \sigma_\alpha^2) \ . \end{gathered} \tag{2}$$

Comparing this to the continuous data situation we see that the distribution assumed for y_{ij}, conditional on the random effects, is a Bernoulli as opposed to a normal distribution, and logit($E[y_{ij}|\alpha_i]$) instead of $E[y_{ij}|\alpha_i]$ is modeled as linear in the fixed and random effects. Otherwise the constructions are the same. The use of the Bernoulli distribution takes care of the connection between mean and variance. The logit transformation maps the interval (0,1) for p_{ij} on to the whole real line, where problems with the upper and lower limits of the p_{ij} disappear. It is then reasonable to assume a normal (or other unbounded) distribution for α_i.

This approach is not without its problems. As we discuss later, the computations for ML (maximum likelihood) or REML (restricted maximum likelihood) for model (2) are quite intensive; much more so than for continuous data.

4. BINARY DATA MODELS

A. Beta-binomial

For a binary variable y, a natural approach to capturing the variability in the mean of y is to model it directly rather than indirectly as in (2). That is, assume a parametric distribution for $p = E(y)$. A logical distribution is the beta distribution, since it is a flexible distribution on the interval (0,1); it is the conjugate prior density from Bayesian analysis and it leads to mathematically tractable results. If y is distributed as a Binomial(n,p) conditional on the value of p and p has a beta distribution with parameters α and β, then the marginal distribution is beta-binomial, i.e.,

$$f(y) = \binom{n}{y} B(\alpha + y,\ n + \beta - y)/B(\alpha,\beta) ,$$

where $B(\alpha,\beta) = \int_0^1 x^{\alpha-1}(1 - x)^{\beta-1} dx$ is the beta function.

A difficulty with this approach arises immediately. How do we allow the values of the parameters α and β to vary in order to form realistic models? Let us consider for continuous data the mixed model with a single fixed effect and nested random effects:

$$\begin{aligned} y_{ijk} = \mu + \eta_i + \nu_{ij} + e_{ijk},\ &\text{where the } \eta_i \text{ are fixed effects ,} \\ \nu_{ij} \sim \text{i.i.d. } &\mathcal{N}(0,\ \sigma_\nu^2) \text{ , and} \\ e_{ijk} \sim \text{i.i.d } &\mathcal{N}(0,\ \sigma_e^2), \text{ independently of the } \nu_{ij} . \end{aligned} \tag{4}$$

This model allows the mean of the y_{ijk} to vary with i and allows the y_{ijk} to be correlated within levels of i and j, i.e., $\text{Corr}(y_{ijk},\ y_{ijk'}) = \sigma_\nu^2/(\sigma_e^2+\sigma_\nu^2)$.

By following a hierarchical specification of a model for the binary data, we can induce a correlation among all the ys that have the same p. Thus, to mimic the correlation structure in model (4), we would use the following specification:

$$\begin{aligned} y_{ijk}|p_{ij} &\sim \text{independent Bernoulli}(p_{ij}), \text{ and} \\ p_{ij} &\sim \text{independent Beta}(\alpha_i,\ \beta_i) , \end{aligned} \tag{5}$$

for $i = 1,2,\cdots,a$, $j=1,2,\cdots,b_i$ and $k=1,2,\cdots,n_{ij}$. This induces a correlation among all the ys within each $(i,\ j)$ combination. Also, since the parameters of the beta distribution depend on i, the mean of the conditional distribution of y_{ijk} given p_{ij} is allowed to vary with i. In this general form (5) also allows the variance of the conditional mean of y_{ijk} to vary from one level of i to the next, which (4) does not.

The likelihood for model (5) takes a relatively simple form. Denoting the number of successes within level (i,j) by $s_{ij} = y_{ij\cdot} = \sum_k y_{ijk}$ the log likelihood can be written as (Williams, 1975)

$$\ell = \sum_i \sum_j \left(\sum_{r=0}^{s_{ij}-1} \log(\mu_i + r\theta_i) + \sum_{r=0}^{n_{ij}-s_{ij}-1} \log(1 - \mu_i + r\theta_i) - \sum_{r=0}^{n_{ij}-1} \log(1 + r\theta_i) \right), \qquad (6)$$

where $\mu_i = \alpha_i/(\alpha_i + \beta_i)$ and $\theta_i = 1/(\alpha_i + \beta_i)$. Closed form maximum likelihood estimators do not exist for this model, so (6) needs to be maximized numerically.

Since the variance of the conditional mean is allowed to have different variances depending on i, we actually get separate estimates of the variances in each of the groups, given by the variances of the beta distribution,

$$\mathrm{Var}(p_{ij}) = \frac{\alpha_i \beta_i}{(\alpha_i + \beta_i)^2(\alpha_i + \beta_i + 1)} .$$

It does not really make sense to try to reparameterize (6) to have a single variance since, as discussed above, the conditional mean and the variance must be related. Noting that

$$\mathrm{Var}(p_{ij}) = \mu_i(1 - \mu_i)\sigma_i , \qquad (7)$$

where $\sigma_i = 1/(\alpha_i + \beta_i + 1)$, Crowder (1978) suggests restricting all the σ_i to have a common value, σ. Note that (7) incorporates the need for the variance to decrease to zero as μ_i approaches zero or one. Also, σ_i is the intraclass correlation coefficient so y_{ijk} and $y_{ijk'}$ are uncorrelated if and only if σ_i is zero. For some situations, σ would be a useful parameter of interest.

The beta-binomial approach is limited in its scope of application. Since we model the correlation by having the correlated Bernoulli variables all selected from a distribution with the same probability of success, we are limited to the type of model (5) where the random effects are nested within the fixed effects. This precludes any sort of regression model which has independent variables specific to each Bernoulli variable though, of course, the μ_i can be modelled as a function of fewer parameters. See Prentice (1986) for an illustration of this latter approach. Also, since we are capturing the variation in the conditional mean with a single distribution, the beta-binomial approach is not amenable to multiple random effects. Thus model (5) is about the most general model possible with this approach.

Rosner (1989) shows how the beta-binomial can be extended to incorporate nested random effects and observation-specific covariates through the use of a conditional model. Tosteson, Rosner and Redline (1991) illustrate this conditional approach.

B. Logit-normal models

A more flexible approach to variance components for binary data is the approach outlined in Section 3. This approach uses a logit function to link the mean of y to the fixed and

random effects and assumes the random effects are normally distributed. Conditional on the random effects u,

$$y_i|u \sim \text{Bernoulli}[E(y_i|u)], \qquad i=1,2,\ldots,n, \qquad \text{logit}[E(y_i|u)] = x_i'\beta + z_i'u$$

and (8)

$$u \sim \mathcal{N}_q(0, D) ,$$

where x_i' and z_i' are the ith rows of X and Z, the model matrices for the fixed and random effects for modelling the vector of logits of the conditional mean of y. This approach has been used by Pierce and Sands (1975), Stiratelli, Laird and Ware (1984), Wong and Mason (1985), Drum (1990), Zeger and Karim (1990) and Karim and Zeger (1990).

C. Probit-normal models

Probit-normal models are a class of models very similar to logit-normal models which arise by replacing the logit function in (8) by the probit function, $\Phi^{-1}(\cdot)$, where $\Phi(\cdot)$ is the standard normal cdf. This gives a model

$$y_i|u \sim \text{Bernoulli}[E(y_i|u)], \qquad i=1,2,\ldots,n, \qquad E[y_i|u] = \Phi(x_i'\beta + z_i'u)$$

and (9)

$$u \sim \mathcal{N}_q(0, D) .$$

This model retains the flexibility of the logit-normal models and has been used by Harville and Mee (1984), Ochi and Prentice (1984), Gilmour, Anderson and Rae (1985), Im and Gianola (1988) and McCulloch (1990).

5. ANALYSIS STRATEGIES

As is evident in the examples in Section 2, analysis goals can differ widely depending on the problem. In situations like Example 3, where only the fixed effects are of interest, it may be tempting to ignore the random effects altogether. In some cases the estimates of the fixed effects are largely the same as those obtained for the marginal distribution in a model incorporating the random effects. However, the standard errors are often grossly mis-estimated. In the salamander data example from McCullagh and Nelder (1989, p. 439) the true standard errors are found to be 40% larger than standard errors calculated from erroneously assuming that the observations are independent, i.e., a logistic analysis. Karim and Zeger (1990), using a Bayesian analysis for the same data set, come to similar conclusions.

Another approach would be to estimate a separate parameter for each level of each random effect. Besides being impractical for a large number of levels (e.g., the aphid data set has 56 clones and 128 sublines), it also leads to erroneous analyses since the observations are

still treated as marginally independent.

For situations like example 4, where only the fixed effects are of interest and the random factor is a nuisance factor and if the data can be segregated into independent blocks (like longitudinal data) then the quasi-likelihood approaches of Liang and Zeger (1986) and Zeger and Liang (1986) are attractive because they do not require the specification of the covariance structure. Such approaches have recently been adapted (Qu and Medendorp, 1991; Qagish, 1991) for estimation of the variance-covariance structure. They are somewhat less attractive in such circumstances since the specification of third and fourth moments is now required.

For analyses which require estimation of the variance of a random effect (examples 1 and 2), or a parametric empirical Bayes approach (example 3), a model which explicitly incorporates random effects is needed. In such a situation, a Bayesian analysis, ML or REML analysis of one of the models discussed earlier is attractive. If the data are longitudinal then quasi-likelihood approaches are possible. References to examples of the application of these models was given in Section 4.

A variety of computational strategies have been employed to address the complexity of ML estimation for variance component models. Numerical integration is straightforward and works for a small (say, up to three) number of nested random effects (Anderson and Aitken, 1985; Im and Gianola, 1988; McCulloch, 1990). Approximations to the likelihood surface by Stiratelli, Laird and Ware (1984) and Harville and Mee (1984)(using posterior modes) and Goutis (1991)(using a saddlepoint approximation) have been utilized. Taylor series approximations to the link function were used by Gilmour, Anderson and Rae (1985), Zeger, Liang and Albert (1988), and Schall (1991). Schall's strategy is an interesting one which adapts a standard iterative technique (Harville, 1973, p. 328) for the linear, normal model. Additionally, simulation approaches have been employed in Zeger and Karim (1990)(Gibbs sampling) and McCulloch (1990)(Monte Carlo EM).

6. A SIMULATION EXPERIMENT

A small scale simulation was performed to compare probit- and logit-normal ML and Schall's (1991) logit-normal approximate technique. Data were generated from the probit-normal model:

$$Y_{ij}|\boldsymbol{u} \sim \text{indep Bernoulli } (p_{ij}) \qquad (i = 1, 2, \cdots, q \qquad j = 1, 2, \cdots, n \qquad N = nq)$$

$$p_{ij} = \Phi(\beta_0 + \beta_1 x_{ij} + u_i)$$

$$u_i \sim iid\ N(0, \sigma_u^2)\ .$$

The x_{ij} were chosen to be equally spaced from -1 to 1 and were ordered as $x_{11} < x_{12} < x_{13} \ldots < x_{21} < \ldots$. β_0 was set equal to zero, β_1 was set to one and q, the number of levels for the random effect, was set to 20. Simulations were run for $\sigma_u^2 = .1$ and .5 and total sample sizes, N, of 200 and 300. Summary statistics for estimation of σ_u^2 are given in Table 1. For $N = 300$ and $\sigma_u^2 = .1$, Figures 1 and 2 show the joint distribution of the probit- and logit-normal MLEs and, for the logit-normal, the MLE and approximate method. Since logit-normal estimates of σ_u are expected to be about $\frac{\pi}{\sqrt{3}} \frac{15}{16} \doteq 1.7$ larger than probit-normal (Johnson and Kotz, 1970, p. 6) we have plotted the reference line $X = \left(\frac{\pi}{\sqrt{3}} \frac{15}{16}\right)^2 Y$. From Figure 1 we see that the probit- and logit-normal estimates are practically identical for each data set and the distribution of $\hat{\sigma}_u^2$ is skewed right. About 6% of the estimates are zero. Among the logit-normal methods, Schall's technique always gave positive values and gave estimates larger than ML. For $\sigma_u^2 = .1$ all methods worked reasonably well, with a slight preference to Schall's method. When $\sigma_u^2 = .5$, Schall's method performed poorly, perhaps because the Taylor series approximation becomes less accurate for large σ_u^2.

Table 1: Estimation of σ_u^2 in a probit-normal model using probit- and logit-normal ML and an approximate logit-normal model. Logit-normal estimates have been divided by $\left(\frac{\pi}{\sqrt{3}} \frac{15}{16}\right)^2$ for comparability. Simulation standard errors are below the estimates.

		$E[\hat{\sigma}_u^2]$ (SE)			$SD(\hat{\sigma}_u^2)$		
σ_u^2	N	Probit	Logit	Logit-Approx	Probit	Logit	Logit-Approx
.1	200	.07	.07	.08	.08	.07	.07
		(.005)	(.005)	(.005)			
	300	.09	.09	.10	.06	.06	.06
		(.005)	(.005)	(.005)			
.5	200	.49	.49	.43	.30	.32	.23
		(.02)	(.02)	(.02)			
	300	.44	.45	.42	.23	.26	.19
		(.02)	(.02)	(.01)			

7. APPLICATION

We now consider the application of the models to the aphid data set. Clearly, an estimate of the variance of a random effect is desired, so a random effects model is needed. Geneticists often hypothesize the existence of an underlying, normally distributed latent variable, so the probit-normal model might be slightly preferred over the logit-normal model. However, as we show, the results from the two models are quite similar. The beta-binomial model is not flexible enough to handle either the multiple random effects or the complicated nature of the fixed effects. We illustrate the application of the probit- and logit-normal models to the aphid data set. We fit a model with random effects of clone and subline nested within clone and fixed effects consisting of a constant term, crop (alfalfa or clover), time (early versus late), plot and a crop by plot interaction. More specifically for the probit-normal model, let i index plot, j index test crop, k index time, l index clone and m index subline, and let

$$Y_{ijklm} = \begin{cases} 1 & \text{if the insect survives to reproduce} \\ 0 & \text{otherwise} \end{cases}$$

and then we assume

$$Y_{ijklm}|\boldsymbol{u} \sim \text{Bernoulli}\,(p_{ijklm}) \ ,$$

where

$$p_{ijklm} = \Phi(\mu_{ij} + \beta_k + u_{1l} + u_{2lm}) \ .$$

We further assume that

$$u_{1l} \sim iid\ N(0, \sigma^2_{\mathsf{CL}}) \quad \text{and}$$

$$u_{2lm} \sim iid\ N(0, \sigma^2_{\mathsf{SL}}) \quad \text{independently}$$

of the u_{1l}. Thus u_{ij} represents the plot and crop effects and the plot by crop interactions, β_k is the time effect and u_{1l} and u_{2lm} are the random effects for, respectively, clone and subline. The logit-normal model we fit is the same except with $\Phi(\cdot)$ replaced by $\dfrac{1}{1+e^{-(\cdot)}}$.

The estimated parameters and log likelihood values are given in Table 2. The estimates of the fixed effects show little change no matter which random effects are included. As expected, the logit-normal fits are nearly identical to the probit-normal with parameter estimates approximately 1.7 times larger. The log likelihood values show that the subline variance component adds very little to the model and can be dropped. A test of the clone variance component can now be performed by referring the difference in $-2*$(log likelihood) to a $\frac{1}{2}\chi^2_1$ distribution ($\frac{1}{2}$ because of boundary effects). The difference is 4.75 giving a p-value less

than .01. However, the estimated within clone correlation on the probit scale is rather small:

$$\text{correlation} = \frac{\hat{\sigma}^2_{CL}}{\hat{\sigma}^2_{CL} + 1}$$

$$= \frac{(.41)^2}{(.41)^2 + 1} = .14 \quad .$$

Table 2: Probit- and logit-normal model estimates and log-likelihoods for the aphid data. Dashes (–) indicate a parameter omitted from the model fit.

Model	Parameter Estimates							−2∗ log likelihood
	μ_{11}	μ_{12}	μ_{21}	μ_{22}	β	σ_{CL}	σ_{SL}	
Probit-	2.0	–.4	1.4	.2	.0	.41	.19	363.33
Normal	2.0	–.4	1.4	.2	.0	.42	–	363.39
	2.0	–.4	1.4	.2	.0	–	.46	365.24
	1.9	–.4	1.2	.2	.0	–	–	368.14
Logit-	3.8	–.7	2.5	.3	.0	.70	.37	363.27
Normal	3.7	–.6	2.4	.3	.0	.73	–	363.37
	3.8	–.7	2.5	.3	.0	–	.81	365.25
	3.5	–.7	2.2	.2	.0	–	–	368.12

8. REFERENCES

Gilmour, A.R., Anderson, R.D., and Rae, A.L. (1985). The analysis of binomial data by a generalized linear mixed model. Biometrika 72: 593-599.

Goutis, Constantinos (1991). Recovering extra-binomial variation. Research Report 79, University College, London, England.

Harville, D.A. (1973). Maximum likelihood approaches to variance component estimation and to related problems. Journal of the American Statistical Association, 72: 320-340.

Harville, D.A., and Mee, R.W. (1984). A mixed model procedure for analyzing ordered categorical data. Biometrics 40: 393-408.

Im, S. and Gianola, D. (1988). Mixed models for binomial data with an application to lamb mortality. Applied Statistics 37: 196-204.

Johnson, N.L. and Kotz, S. (1970). Continuous Univatiate Distributions –2. Wiley, New York.

Karim, M.R., and Zeger, S.L. (1990). Generalized linear models with random effects; salamander mating revisited. Johns Hopkins University, Department of Biostatistics Technical Report 722.

Liang, K.-Y. and Zeger, S.L. (1986). Longitudinal data analysis using generalized linear models. Biometrika 73: 13-22.

McCullagh, P. and Nelder, J.A. (1989). Generalized linear models. (Second Edition). Chapman and Hall, London.

McCulloch, C.E. (1990). Maximum likelihood variance components estimation for binary data. Biometrics Unit, Cornell University Technical Report BU-1037-MA.

Ochi, Y. and Prentice, R.L. (1984). Likelihood inference in a correlated probit regression model. Biometrika 71: 531-543.

Pierce, D.A. and Sands, B.R. (1975). Extra-Bernoulli variation in binary data. Technical Report 46. Department of Statistics, Oregon State University.

Prentice, R.L. (1986). Binary regression using an extended beta-binomial distribution, with discussion of correlation induced by covariate measurement errors. J. of the Amer. Stat. Assoc. 81: 321-327.

Qaqish, B.F. (1991). Computing for generalized estimating equations. Presentation at the Biometric Society Meeting, Houston, Texas, March 24-27.

Qu, Y. and Medendorp, S.V. (1991). Regression models for correlated ordinal data. Presentation at the Biometric Society Meeting, Houston, Texas, March 24-27.

Rosner, B. (1989). Multivariate methods for clustered binary data with more than one level of nesting. J. of the Amer. Stat. Assoc. 84: 373-380.

Schall, R. (1991). Estimation in generalized linear models with random effects. Technical Report, Division of Biometry, University of the Orange Free State, Tygerberg, South Africa.

Stiratelli, R., Laird, N.M., and Ware, J.H. (1984). Random-effects models for serial observations with binary response. Biometrics 40: 961-971.

Tosteson, T.B., Rosner, B. and Redline, S. (1991). Logistic regression for clustered binary data in proband studies with application to familial aggregation in sleep disorders. To appear in Biometrics.

Weil, C.S. (1970). Selection of the valid number of sampling units and consideration of their combination in toxicological studies involving reproduction, teratogenesis or carcinogensis. Food and Cosmetic Toxicology 8: 177-182.

Williams, D.A. (1975). The analysis of binary responses from toxicological experiments involving reproduction and teratogenicity. Biometrics 31: 949-952.

Wong, G.Y. and Mason, W.M. (1985). The heirarchical logistic regression model for multilevel analysis. Journal of the American Statistical Association. 80: 513-524.

Zeger, S.L. and Karim, M.R. (1990). Generalized linear models with random effects; a Gibb's sampling approach. To appear in Journal of the American Statistical Association.

Zeger, S.L. and Liang, K.-Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. Biometrics 42: 121-130.

Figure 1: Comparison of estimates of σ_u^2 by probit- and logit-normal ML probit-normal model. Logit-normal estimates have been divided by $\left(\frac{\pi}{\sqrt{3}}\ \frac{15}{16}\right)^2$ for comparability. Sample size n = 300, $\sigma_u^2 = .1$.

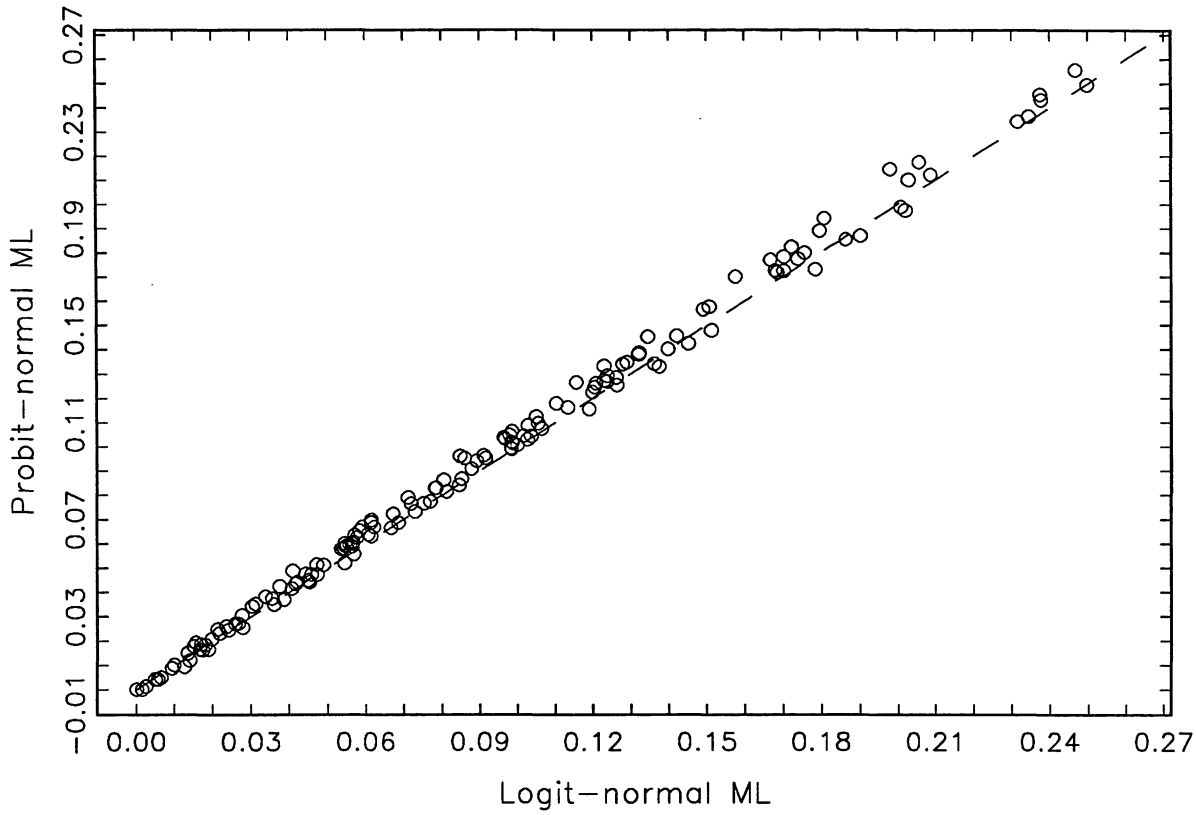

Figure 2: Comparison of estimates of σ_u^2 by an approximate logit-normal method and logit-normal ML probit-normal model. Estimates have been divided by $\left(\frac{\pi}{\sqrt{3}}\ \frac{15}{16}\right)^2$ for comparability to the probit-normal model. Sample size n = 300, $\sigma_u^2 = .1$.